NATIONAL BUREAU OF STANDARDS REPORT

6595

A NEW APPROACH

to the

MECHANICAL SYNTACTIC ANALYSIS OF RUSSIAN

Mechanical Translation Group
U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS
Applied Mathematics Division

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

THE NATIONAL BUREAU OF STANDARDS

Functions and Activities

The functions of the National Bureau of Standards are set forth in the Act of Congress, March 3, 1901, as amended by Congress in Public Law 619, 1950. These include the development and maintenance of the national standards of measurement and the provision of means and methods for making measurements consistent with these standards; the determination of physical constants and properties of materials; the development of methods and instruments for testing materials, devices, and structures; advisory services to government agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; and the development of standard practices, codes, and specifications. The work includes basic and applied research, development, engineering, instrumentation, testing, evaluation, calibration services, and various consultation and information services. Research projects are also performed for other government agencies when the work relates to and supplements the basic program of the Bureau or when the Bureau's unique competence is required. The scope of activities is suggested by the listing of divisions and sections on the inside of the back cover.

Publications

The results of the Bureau's work take the form of either actual equipment and devices or published papers. These papers appear either in the Bureau's own series of publications or in the journals of professional and scientific societies. The Bureau itself publishes three periodicals available from the Government Printing Office: The Journal of Research, published in four separate sections, presents complete scientific and technical papers; the Technical News Bulletin presents summary and preliminary reports on work in progress; and Basic Radio Propagation Predictions provides data for determining the best frequencies to use for radio communications throughout the world. There are also five series of nonperiodical publications: Monographs, Applied Mathematics Series, Handbooks, Miscellaneous Publications, and Technical Notes.

Information on the Bureau's publications can be found in NBS Circular 460, Publications of the National Bureau of Standards ($1.25) and its Supplement ($1.50), available from the Superintendent of Documents, Government Printing Office, Washington 25, D.C.

NATIONAL BUREAU OF STANDARDS REPORT

NBS PROJECT
1102-40-11513

November 10, 1959.

NBS REPORT
6595

A NEW APPROACH
to the
MECHANICAL SYNTACTIC ANALYSIS OF RUSSIAN

Mechanical Translation Group
U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS
Applied Mathematics Division

IMPORTANT NOTICE

NATIONAL BUREAU OF STAN
intended for use within the Go
to additional evaluation and revi
listing of this Report, either in
the Office of the Director, Natio
however, by the Government ag
to reproduce additional copies

Approved for public release by the director of the National Institute of Standards and Technology (NIST) on October 9, 2015

ress accounting documents
ly published it is subjected
roduction, or open-literature
is obtained in writing from
h permission is not needed,
pared if that agency wishes

NBS

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

A NEW APPROACH

to the

MECHANICAL SYNTACTIC ANALYSIS OF RUSSIAN*

Mechanical Translation Group
U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS
Applied Mathematics Division

* This work was sponsored by the Office of Ordnance Research, Department of the Army.

NBS 156
Manuscript
Paper

GPO

16—11141-1

A NEW APPROACH TO THE MECHANICAL SYNTACTIC ANALYSIS OF RUSSIAN

INTRODUCTION

From the moment that a writer conceives an idea which he desires to communicate to his fellow men, sizable stumbling blocks are strewn in the path of the future translator. For the ability to shape one's thought clearly, or even completely, is not granted to many; rarer still is the gift of expressing the thought--precisely, concisely, unambiguously--in the form of words. There is no guarantee, therefore, that the author's written text is a reliable image of his original idea.

Furnished with this more or less distorted record, the translator is expected to perform a number of amazing feats. In the first place, he has to discern--through the dim mist of the source language--the writer's precise intention. This requires not only a perfect knowledge of both the source language and the subject matter treated in the text, but also the mental skills customarily exercised by the professional sleuth. In addition, these newly reconstructed ideas must be rendered into a target language which is so limpid, so unequivocal, so faithful to the source--as to convey, to every reader of the translator's product, the exact <u>intention</u> of the original foreign text!

Small wonder, then, that a fabulous achievement like Fitzgerald's translation of the Rubaiyat is regarded in the nature of a miracle. For the general case, it would seem that characterizing a sample of the translator's art as a <u>good</u> translation is akin to characterizing a case

NBS 156
Manuscript
Paper

of mayhem as a good crime: in both instances the adjective is incongruous.

If, as a crowning handicap, we are asked to replace the vast capacity of the human brain by the paltry contents of an electronic contraption, the absurdity of aiming at anything higher than a crude practical translation becomes eminently patent.

Perhaps we are belaboring this point; we do so to avoid later arguments about the "quality" of our work. If, for example, a translated article enables a scientist to reproduce an experiment described in a source paper and to obtain the same results,--such a translation may be regarded as a practical one. Perhaps the translation is not couched in elegant terms; here and there several alternative meanings are given for a target word; a word or two may appear as a mere transliteration of original source words. Nevertheless, this translation has served its main purpose: a scholar in one land can follow the work of his colleague in another.

This limited scope has been set for us by our own as well as the machine's deficiencies. The heartbreaking problem which we face in mechanical translation is how to use the machine's considerable speed to overcome its lack of human cognizance gained during a lifetime of experience. We do not yet really understand how the human mind associates ideas at an almost infinite rate of speed; for example, how does it differentiate instantaneously between the two meanings of "calculus" in the following sentences: (1) The surgeon removed the staghorn calculus from the patient's kidney, and (2) The professor announced a new course

NBS 156
Manuscript
Paper

GPO

16—11151-1

in advanced calculus. And yet, a scheme for discerning such differences is precisely what we must impart to the machine.

Even if there now existed a completely satisfactory method for machine translation, today's machines would not be adequate tools for its implementation. They lack automatic transformers of printed text into coded signals, and their external storage devices are not up to the mark.

Before coming to grips with the mechanical translation problem, we investigated the types of difficulties we might encounter. We found that they fall into ten groups; so far, we have been able to cope--more or less successfully--with only the first five, which depend mainly on syntactic analysis. Some thought has been given to the far more difficult points involving semantic considerations, but the short time spent in this area has not allowed us to transform the mathematical "existence solutions" into practical machine application. Thus, discussion of semantic problems is deferred. In this paper we are concerned mainly with syntactic analysis.

THE GLOSSARY

One of the indispensable accessories of MT is the construction of a source-to-target Glossary.[1/] The conventional publications would not suffice for MT, because their authors presuppose, on the part of the prospective user, (1) a wide acquaintance with the basic principles of the source language, (2) an excellent knowledge of the target language, and (3) a considerable familiarity with the terminologies--in both languages--relating to the special subject of the source text. These assumptions are hardly justified even in the case of the professional translator. It follows that a glossary, designed for use with an electronic processor, must embody an immense amount of information in addition to the material culled from the best existing dictionaries. But there is a limit to the amount of data that can be handled by even the most advanced type of electronic processor, if MT is to be at all expedient. It is imperative, therefore, that utmost care be used to select (1) the absolutely minimum quantity of information which would suffice for our needs, (2) the most economical (space and time-saving) form for representing it, and (3) the most suitable external media for its storage and retrieval.

1/ The List of Terms and List of Symbols at the end of the paper may enable the reader to identify unfamiliar expressions. Technical words to be found therein are capitalized when first encountered in the text.

NBS 156
Manuscript
Paper

GPO

16--11141-1

Of far greater concern is the fact that no genius exists today who is completely aware of the mental processes involved in the performance of the translation task. Yet a routine, duplicating these processes, must be prepared for insertion into the machine's memory. Unfortunately, the form of the glossary depends completely upon, and varies constantly with, the particular translation scheme which is being developed by its human originator. We would not venture to predict the date when our own glossary might assume its final--or even "passable"--shape. We are constrained, for the present, to use a small sample glossary, sufficient for trial runs on the computer. It is stored in the external memory and is arranged in groups, each of which lists the Satellites of a source Pseudo-root. Each satellite is an entry corresponding to a source Stem which contains the pseudo-root in question. The temporary form, which each Glossary Entry has assumed so far, consists of the following items:

1. The Source Transform, which is a greatly contracted form of the original source stem.

2. Morphological information, designed to aid in the syntactical analysis of each sentence, as illustrated in Section B of Part II.

3. Predictions regarding future Occurrences. For instance, the Russian verb with stem СЛУЖ is marked as frequently followed by an indirect object in the dative case and/or a complement in the instrumental; also sometimes by a verb in the infinitive.

4. One or more target correspondents (T) to the source stem.

(It is planned to expand this information to include diacritical material designed to aid in the semantic analysis of the sentence.)

PART I.

Our program is being coded in two parts. Of these only the first, which consists of two sections, has been completed and tested.

Section A.

The aim of this section is to investigate the nature of each occurrence in a sentence and, for the case when the occurrence is a word, to perform a glossary look-up.

When an occurrence in a given Russian text is read into the machine--and we have reason to hope that this will be accomplished eventually by a fully automatic device--this source material is subjected to the following treatment within the computer.

1. An Identification Tag (t) is appended to the occurrence to indicate the page, sentence, and ordinal number. Its characters are counted and examined for indications anent its physical make-up. For instance, the machine examines whether the occurrence is a word or, perhaps, a punctuation mark, formula, etc. If a word, it notes whether it starts with a capital or is an initial, whether it contains any indication of foreign origin. This orthographical material will be augmented and revised in succeeding steps to form General Specifications (GS). It is recorded in the internal memory space S_t, allotted to the occurrence t.

NBS 156 Manuscript Paper

GPO

10--11141-1

2. If the current occurrence is not a word, this fact is signalized in the Profile Skeleton (PS) which will eventually be expanded to serve as a rough outline of the clause formation of the source sentence to which the occurrence belongs. If, moreover, the occurrence is identified as a period, a subroutine is consulted to determine whether this punctuation marks the end of the sentence. If such be the case, this fact is indicated in the profile skeleton, and the sentence number is raised for storage in the succeeding tag numbers, t.

3. If the given occurrence is a word, a search is made in a Special List of frequently used words. If the word is found in the special list, the diacritical material accompanying it may show that it could be the leading word of one or more idioms. In that case, the requisite number of successive source occurrances will be compared to each of the indicated idioms, and when agreement is found, the entire source idiom is replaced by the corresponding material and is thereafter treated as a single occurrence.

4. If the word is not found in the above list, it is decomposed into its Pseudo-prefixes, pseudo-root[1/] (or roots), Pseudo-suffixes, and Source Ending[1/] by means of corresponding Lists stored in the internal memory.

1/ The pseudo-root and true source ending are determined by a rather complicated iterative scheme.

The ending is replaced by the address β, found alongside its listed counterpart. It is stored in S_t and will be used in Part II.

Each pseudo-prefix and pseudo-suffix (if any) is replaced by a single character, consisting of 6 bits, and the combination of these characters (probably no more than 8) constitutes the transform (Δ) of the original source word; y and z, the number of pseudo-prefixes and pseudo-suffixes, as well as Δ, are stored in S_t.

The remaining portion of the current word, constituting the pseudo-root, may have no characters at all. The glossary contains a group of satellites for a null pseudo-root. Its Extended Address, α_o, is used to represent it in the next step.

If the pseudo-root contains at least one character, it may not have been found in the list of pseudo-roots. In that case, the transliteration subroutine dictates the form of the correspondent to be stored in the normal position of the target T for the final printout. A suitable Signal of Peculiarity (δ) is stored in GS. The Correspondence Flag (c) in GS is set to zero.

If the pseudo-root has been located in the list, its counterpart is accompanied by an extended address, α, indicating where its group of satellites starts in the externally stored glossary.

5. The extended address α, accompanied by the identification tag t, is intersorted with similar combinations, corresponding to the previously processed source words, in the Sorting File.

6. When all the internal space alloted for the sorting file is

filled, a search is made throughout the entire glossary for the indicated entries. Since the time for such a transit throughout the glossary is formidable, and remains practically constant irrespective of the number of words to be looked up, it is obvious that an appreciable increase in internal storage space would result in a corresponding reduction in the look-up time per word. However, considering the high cost of internal storage devices, it might be more expedient to utilize inexpensive non-erasible external storage media with suitable buffering devices which allow for the simultaneous retrieval of information along several busses.

7. When the extended address α attached to t is reached during transit of the glossary, the routine searches for the entry corresponding to the y.z.Δ of the occurrence t. The correspondence flag c is set to 1 or 0 in GS, according to whether the search has been successful or not. In the latter case, the pertinent peculiarity signal is stored in GS and the tag t is placed in the normal position of the target T for final printout.

Illustration 1.

As an example of the performance of this section of the program, we offer the text word РАСПОЛОЖЕНИЕ. Suppose this word occurs as the 7th word of the 4th sentence on page 1. The corresponding symbol for t is 1.4.7. The occurrence is examined and found to be a word (not a punctuation mark etc.) composed of 12 letters. The Word Flag (w) in GS would be set to 1.

The machine determines that no such word appears in the special list of frequently used words. The occurrence is therefore examined for pseudo-prefixes. In this case, the combinations РАС and ПО happen to be true prefixes. By referring to the stored list of pseudo-prefixes, the routine would replace РАС by the letter V, and ПО by the letter R. Unable to discover more prefixes, the routine would isolate the ending ИЕ. Suppose that the list of endings indicates that information on this ending is stored in internal memory beginning at address 357; the machine then sets $\beta = 357$. The routine would proceed to identify ЕН as a suffix and replace it by the letter K. Finding no more pseudo-suffixes, the routine would store in $S_{1.4.7}$ the numerals 2 and 1, to indicate the number of prefixes and suffixes y and z; these would be followed by the transform Δ, which is VRK. The machine would then enter the subroutine for identifying the pseudo-root. In the present case, no difficulties would be encountered, as ЛОЖ would be located at once in the list of pseudo-roots. In actual practice, a number of complications may arise. The given word may contain a polyroot; or what we assumed to be an ending may actually be part of the pseudo-root; or we may not be able to locate the root at all. The sub-routine takes note of all these possibilities.

The root ЛОЖ is replaced by α which would be, say, 2.47.3097, if the first member in the group of this root's satellites has the position number 3097 in the 47th block on the 2nd tape. To α we attach the tag t and intersort the result with the other contents of

the sorting file. The entry in the internal memory, corresponding to the occurrence РАСПОЛОЖЕНИЕ, now has the two forms

Storage	GS	β	y.z	Δ
$S_{1.4.7}$	Orthographic description	357	2.1	VRK

	α	t
Sorting File	2.47.3097	1.4.7

After a specified number of successive occurrences have been analyzed in this way, a transit will be made through the glossary. When the position 3097 of the 47th block on the 2nd tape is reached, the machine will locate and extract all the material corresponding to 2.1. VRK, i.e. all the information pertinent to the stem РАСПОЛОЖЕН. In GS, the correspondence flag c would be set to 1 to indicate that the search had been successful.

Section B.

In this section we examine each word-occurrence of a sentence with two aims in view:

1. To assign to it all possible grammatical interpretations, which we call Temporary Choices, TC_j. These are arranged roughly in order of most probable appearance; j indicates the serial number.

Information common to all TC_j is labeled with $j = 0$.

2. To indicate its significance in the profile skeleton.

To accomplish the first aim we distinguish three types of words:

a. If a source word is found in the special list of frequently used words, its various target meanings serve as the TC_j.

b. For a word whose transform is found in the glossary, the TC_j are obtained by finding the common intersection between the possibilities given by its ending in the Table of Endings and those given by the morphological information of the stem's glossary material.

c. When a source word is represented merely by its transliteration, the TC_j must be made on the basis of its ending (and, possibly, its suffixes) only.

As regards the second aim, the TC_j which accompany a current word may reveal that it could be a possible indicator of a main clause, or subordinate clause, or a phrase. If such is the case, an appropriate signal is added to the profile skeleton, in which the nature of the non-word occurrences has previously been stored. The profile skeleton will be subjected to a crude analysis in Section A of Part II.

Illustration 2.

Let us use again the word РАСПОЛОЖЕНИЕ, belonging under the heading 2b above. The glossary's morphological information indicates that its stem, РАСПОЛОЖЕН, could represent either

NBS 156 Manuscript Paper

GPO

16—11141-1

1. An inanimate neuter noun, belonging to a declension class which is identified by the ending ИЕ in the nominative singular; or

2. An adjective, of verbal origin, belonging to a declension class which is identified by the ending ЫЙ in the masculine nominative singular.

This material, used in conjunction with the information listed for the ending ИЕ, leads the machine to eliminate the second possibility given by the glossary and to list the following two temporary choices:

TC_0	Noun, inanimate, neuter	
TC_1		nominative, singular
TC_2		accusative, singular.

PART II

Part II of the projected scheme, now in process of being programmed, has the purpose of analyzing the syntactical structure of each source sentence and of constructing a corresponding target sentence. While Part I works on at least several hundred source words in one pass--the precise number of such words is determined by the internal memory capacity of the machine--Part II, which is made up of three sections, works on one sentence at a time.

Section A determines, as far as possible at this stage, the clausal and phrasal structure of the sentence. Section B is an iteration scheme for examining syntactical relations among the Strings of a sentence. It processes each string in turn from the beginning to the end of each sentence, repeats this process if necessary and decides whether a translation has been effected. Thereafter Section C takes over, composes a target sentence, and prints it out.

Types of Difficulties.

We shall list, in order of increasing complexity, the ten difficulties which obstruct our path toward such a goal:

1. The stem of a source word is not listed in our glossary. This will occur quite often in our translation scheme, as we intend to omit from the glossary the majority of non-slavic stems.

2. The target sentence requires the insertion of key English words, which are not needed for grammatical completeness of the source sentence. For instance, the complete Russian sentence: ОН БЕДНЫЙ (literally: He poor.) must be translated as: He (is) (a) poor (man).

IBM 156 Manuscript Paper

GPO

16—11141-1

3. The source sentence contains well-known idiomatic expressions.

4. The occurrences of a source sentence do not appear in the conventional order. Sober writing, without color or emphasis, employs few inversions. Our method, which consists of predicting each occurrence on the basis of the preceeding ones, works quite well in that case. But such orderliness cannot be expected to hold for long stretches of the text.

5. The source sentence contains more than one clause.

6. Corresponding to an occurrence in the source sentence, more than one target word is listed in the glossary. Polysemy is, of course, recognized as a most formidable obstacle to faithful translation, whether human or mechanical. Hilarious (or heartbreaking, depending on your point of view) "malaprops" can be cited by the score to uphold the conviction of many linguists that the MT task is a hopeless one. Our faith in the inventiveness of the human brain makes us reject such gloomy forebodings.

7. The source sentence is grammatically incomplete. Such a situation is frequently the result of carrying on the thought from one or more previous sentences. To succeed, any MT scheme will have to be able to transcend the boundaries of a sentence (or a paragraph, or a section).

8. The source sentence contains ambiguous symbols. Since we are planning to confine our efforts to mathematical texts, such occurrences will be legion.

9. The syntactic integration of the source sentence results in an ambiguity. It is often of a type that could be resolved by semantic considerations; but sometimes, it is inherent and thus not removable by any process.

10. A combination of difficulties are listed in this category. They are quite annoying but fortunately rare: misprints; grammatical errors; localisms; peculiar nuances; comments based upon the sound (or the spelling) of source occurrences, such as puns whose sense it is impossible to render into the target language.

We have thus grouped Russian sentences into 2^{10}, i.e. 1024, types. A sentence possessing none of the ten difficulties would be represented by type number $00000\ 00000_2$ whereas--at the other end--a sentence exhibiting all the difficulties would belong to type $11111\ 11111_2 = 1023_{10}$.

Our scheme is able to cope successfully--we believe--with the first five types of difficulties, which involve only monosemantic occurrences, or at most idiomatic expressions. We can thus handle 32 types of sentences ranging in type number from $00000\ 00000_2$ to $00000\ 11111_2$.

<u>Section A.</u>

In both sections of Part I we kept up, for each source sentence, a profile skeleton which consists of a set of signals denoting to which special class (if any) each occurrence belongs. This tentative outline serves to indicate where the clauses and phrases of the sentence might have their inception. The routine in the present section

carries out an iterative process which aims to set rough limits to these ranges, based upon the position in the sentence of its (1) punctuation marks, (2) conjunctions, (3) actual, or possible, starters of main clauses, (4) actual, or possible, starters of subordinate clauses, (5) actual, or possible, predicates for each clause, and (6) actual, or possible, phrase starters.

As a result of this iterative scheme, the profile skeleton PS is replaced by a Temporary Profile (TP), in which each occurrence is associated with four designators:

1. Its clause number (C),

2. A Status Flag (v) to indicate whether the predicate of the clause has or has not occurred,

3. Its phrase number (P), and

4. A Backward Flag (b) to indicate a particular manner in which the string is to be handled during the process of syntactic integration.

In the event that the routine does not succeed in determining a clause or phrase number, it will insert a Signal of Uncertainty (X), which the routine in Section B will attempt to resolve.

Section B.

At the conclusion of the preceeding section, each source occurrence has been replaced by a string of information which will expand as we progress in the integration scheme. The string, at this point, contains several sets of data:

A set of general specifications, GS, consisting of

a. a word flag, w, indicating whether the occurrence was or was not a Word-utterance (W).

b. a correspondence flag, c, indicating whether or not the occurrence (or its transform) was located in the storage.

c. a peculiarity signal, δ, pointing out any significant feature of the occurrence.

2. A set of four designators, belonging to the temporary profile, TP.

3. If the occurrence was a W, its string will have in addition

a. a set of temporary choices, TC_j, giving all possible grammar interpretations of the source word.

b. a set of target correspondents, T, if the word (or its transform) has been located in the memory; otherwise the correspondent will be either

1) the transliteration of all, (or part) of the word-utterance, if its pseudo-root is not listed; or else

2) the identification t, if its transform is not in the glossary.

c. a set of Glossary Predictions (GP) retrieved from the memory if such exist, each consisting of

1) a Grammar Essential (GE), indicating the predicted type of agreement with a temporary choice.

2) a Signal of Urgency (u) indicating the probability of fulfillment.

In addition to the above items, there may be available at any stage of the iterative process in Section B the following information, which has been generated during the preceding portion of this Section and incorporated within the strings which cause it to be generated.

1. Foresight Predictions (FP). Expectations for future strings, based on past occurrences; e.g. a direct object is governed by a transitive verb. A foresight prediction contains at least three specifications:

a. Serial number, k, to distinguish the different foresights generated by the same string.

b. Urgency Code (U), designating the degree of necessity--or the proximity--of the expected string, (e.g. a code of 1 indicates: next occurrence or not at all).

c. Sentence Element (SE), such as Subject, Predicate, Complement, etc.

In addition to the above items, which are always present, a foresight prediction may contain data, in the form of

d. Morphological Specifications (MS) regarding animation, gender, number, etc.

e. An Insert Flag (e) to indicate whether or not an English preposition is to be inserted before the target correspondent, T.

2. Hindsight (H_1) regarding troublesome strings. When a Predictable Choice does not agree with any of the previous FP,

NBS 156
Manuscript
Paper

GPO
16—11141-2

Hindsight Entries about this Unexpected Choice are stored together with a Chain Flag (f) in H_1, to be considered with subsequent strings. Such apparent inconsistencies must all be resolved at the conclusion of the sentence, as a necessary (but not sufficient) criterion of successful syntactical integration. Here, too, are stored queries about strings whose syntax is questionable, even though they seemingly fulfill previous predictions. Entries in H_1 concerning these Doubtful Choices are not flagged.

3. Hindsight (H_2) regarding predicted alternate temporary choices. It may happen that more than one of the temporary choices TC_j agree with previously made predictions. In this case, one is selected as a link in the sentence structure and the others are stored for future consideration in the current (and subsequent) iterations.

4. Hindsight (H_3) regarding the remaining unpredicted temporary choices TC_j. These are "pigeonholed" for possible use in subsequent iterations.

5. Chain number (L). Whenever the machine, in proceeding through a sentence, encounters a string which it is unable to link with any previous predictions, it starts a new Chain. There exist, however, five types of Unpredictable Choices which do not cause the chain number to be raised. They represent (a) punctuation marks, (b) conjunctions, (c) adverbs, (d) particles, and (e) prepositions.

The Routine of Section B begins with the following steps:

1. All the hindsight entries, left in storage from the previous

sentence, are cleared out.

2. The chain number L is set to 1.

3. The following two predictions, for the main clause, are stored as foresights:

k.U.SE

1.7.Subject

2.7.Predicate

where k is the serial number within the string; U is the urgency code (7 indicates the highest); and SE is the sentence element of the prediction.

We now attempt to determine the syntactic sentence structure by observing the following routine for each string. (The letter q will indicate the current String number; Q will denote this running coordinate as it ranges from 1 to q;) K and J will denote, respectively, the k and j within the string Q.

1. The routine examines the unfulfilled FP_{QK} within the current clause or phrase, in decreasing order of Q and increasing order of K. Each of them is tested for agreement with any of the TC_j. The first TC which fits an FP is taken as the Selected Choice (SC) for this iteration. The successful FP is deleted. If there are several TC_j and none of them fit any FP_{QK}, the hindsight information is examined for possible clues regarding the selection of a TC_j to act as the SC. If no clue is found, TC_1 becomes the SC. If, however, the string was marked by a backward flag b, the examination of foresight predictions

NBS 156
Manuscript
Paper

GPO
16—11141-4

is omitted. In this case the routine examines--in reverse order--the previous selected choices, SC, for agreement with TC_j. If the string is of the unpredictable type, TC_1 is taken as the SC.

2. The selected choice is indicated by Q.K.j., where Q is the number of the string where the successful prediction (if any) was made and K is the serial number of that prediction. If there is no such prediction for SC, both Q and K are designated as 0. The letter j, of course, represents the serial number of the chosen TC in the current string.

3. The chain number L is left unchanged, if the string has been predicted or is of the unpredictable type; otherwise L is raised by unity.

4. The designators C, v, and P of the temporary profile TP are revised--in the light of the SC--to form the Selected Profile (SP). The status flag v furnishes clues for the subsequent revision of the clause number C, and the syntactical integration determines the bounds of each phrase.

5. New predictions for the foresights are culled from three sources:

a. The temporary profile, TP, of the next string. If the TP indicates that a new clause is starting, the predictions of a new subject and predicate are entered as foresights.

b. The main routine. This may yield predictions of a

general nature on the basis of the SC. For example, if the SC is a noun, one such prediction states that the noun might be followed by a complement in the genitive case. If the SC is the subject, we examine whether the predicate has been found previously; if not, we add to the FP of the predicate the information that it must agree with the subject in person, number, gender, etc. Similarly, if the SC is the predicate, the FP of the subject--if unfulfilled--is amplified.

c. The glossary predictions, GP, accompanying the chosen TC. Such predictions, if any, would arise from the peculiar nature of the original occurrence. For instance, a particular verb may govern the dative case.

6. The predictions yielded by a string are appraised against the entries previously placed in hindsight, in order to ascertain whether the former throw any light upon the difficulties and conflicts represented by the latter. If a partial explanation is obtained, a suitable notation is made alongside the corresponding entry. Whenever such an entry is completely explained away, it is deleted. If such a deletion takes place in H_1, the chain number L is reduced by one, provided the entry bears the chain flag f. Sometimes, a rearrangement in order of the strings is indicated, as a result of the above appraisal.

7. The SC may indicate that a key target word, such as a noun or a verb, has not been explicitly stated in the source sentence. If

such be the case, the routine determines the required Target Insert (TI) and constructs a corresponding New String. On the other hand, the SC may dictate the suppression of a target correspondent(s).

8. A target order number R is assigned to the string, to indicate the arrangement of occurrences in the target language. In general, the R's are consecutive. If, however, the appraisal in Step 6 calls for a rearrangement of strings, or if Step 7 resulted in the insertion of a new string (or the suppression of an Old String)--the affected R's are renumbered in accordance with the desired sequence. Pretarget Inserts (PI), such as prepositions and articles, are not assigned an R. Their handling will be discussed in Section C.

9. The TC_j which do not become the SC may, under certain circumstances, be disregarded. In the cases where the routine directs the machine to retain them, they are entered into hindsight H_2 or H_3, according to whether they do or do not agree with any FP.

10. If the chain number L was raised in Step 5, an appropriate query is entered into hindsight H_1 with a chain flag f. If the SC is a doubtful choice, suitable queries--unaccompanied by the chain flag--are also entered into H_1.

When the end of the sentence is reached, we need not embark upon another iteration if (1) the foresights do not contain unfulfilled predictions of urgency 6 and 7, and (2) the chain number is 1. (In that case H_1 should be clear of flagged entries.)

In this event, the selected choices for all strings are considered

as Final Choices (FC) and the routine proceeds to Section C. If however, another iteration is indicated, it investigates the H_2 information where resolution signals were placed during the previous iteration whenever some partial light was thrown upon any of its entries. As a result, one of the former selected choices is replaced by a more promising one, and the effect of that change is investigated. It is obvious that, if the number of unresolved entries in H_2 is high, it would be prohibitive to pursue all the possible combinations of selected choices. We therefore set a limit to the number of iterations we allow the machine to execute. In the unlikely event that all the possibilities inherent in the H_2 entries have been exhausted, the H_3 entries are attacked in the same manner.

Failure is conceded when the number of iterations already performed has reached the limit we had set for ourselves, or when the current set of selected choices repeats any of the previous sets (which are stored in the internal memory.) In that case, the routine records a failure signal and indications of the types of errors encountered, to be printed out at the conclusion of Section C.

Section C.

This section is devoted to the construction and printing of the target sentence.

1. The target correspondents listed with the final choices are arranged in the sequence, given by R.

2. A subroutine supplies new pretarget inserts PI, in addition to

those supplied by the foresights. These may be either English articles or prepositions. The set of PI (if any) are inserted in front of the proper correspondent for eventual printout.

3. A second subroutine affixes Pidgin Endings (E) to target correspondents whenever needed. (To conserve precious internal space, we regard--for the present--all English targets as grammatically regular. Thus the plural of "foot" will appear as "foot-s".)

4. A count is made of all unresolved hindsight entries.

5. The resulting information is printed out. All inserts, whether PI or TI are printed in parentheses. Words for which there are no target correspondents are enclosed in brackets. They may appear as some combination of the following word-sections:

a. a translated initial prefix

b. a transliterated full or partial stem

c. a transliterated full or partial word.

If the iterative routine failed to satisfy our criteria, this fact would be indicated by the failure signal and by the notations of the error types encountered. On the other hand, the satisfaction of the criteria is no guarantee that the result is a faithful translation, unless all three hindsights are clear and all occurrences are monosemantic. Since such eventualities will be extremely rare, we shall regard the tallies for the hindsight entries and the multiplicity of the printed meanings as a measure of the "goodness of fit" of our version.

S 156
uscript
aper

GPO
16—11141-8

Illustration 3.

The chart given below outlines the syntactic integration of a sentence possessing the five types of difficulty which our routine is able to handle with some degree of success. On the other hand, it contains a number of polysemantic words, of which only a few can be resolved at present. For the remaining polysemantic words, we are forced to print out all the meanings contained in our glossary.

The chart is followed by a number of notes, intended to clarify some of its entries. It is hoped that the List of Terms and List of Symbols may further aid the reader.

NBS 156
Manuscript
Paper

GPO

16—11141-1

0
1
2
3
4
5

PART I								PART II																
	Source Occur'ce		Gen. Spcfcs. (GS)	Temp. Chs. (TC_j)	Morph. Spcfcs. (MS)	Glss. Preds (GP) GE & u	PS	Sec. A. TP	Forcs't Prds. (FP_{qk})		Morph. Sps. (MS)		Section B Sel. Ch. (SC) FP		L	SP	Hindsight H_1 Entry	H_1 Resol.	H_2 Entry	H_2 Res	H_3	Section C Translation		
Q		∅	w | c | S	j	Ps | A | D | ts | C | Y | Z	g | a | d | i | x | n		C | v | P | b	r	k | U | SE	V | A | G | Y | Z	e	Q | K | j	B		C | v | P	f | j | ES	h | Q | K | J	Q | K | j | ES	h | Q	j	R	PI	T	E
1	до- каз- а- ть		1 1 Csp	1	Vb if	7 1 1 1	vi	1 0 0	1 (2)	1 7 Sbj 2 7 Pdc			1 1 1		1	1 0 0	SE Sbj 1 Y m	- 6 1 - 8 2 - 4 3 - 10 3 - 26 1				1	(To)	de- mon- str- ate-
2	без	1	1 1	1	Pp	3	po	1 0 1	24 6 6 24 9	1 7 Pdc 2 7 Cpl 3 6 Cls 4 3 Cpl 5 3 Cpl	vf N 3 s a a d i	 1 1	0 0 1		1	1 0 1						2		with- out
3	по-с- ред- ств- а		1 1	0 1 2 3	Nc in N g s n p a p	1 1		1 0 1	3	1 7 Cpl	g		3 1 1		1	1 0 1						3	(the)	means
4	трансц- ендент- ал-ь- ого		1 0 Nar	0 1 2	Aj ps (F) g s am M a s			1 0 X	4 4 6	1 1 Cpl 2 0 Mfr 3 3 Cpl	g im N g s m	1	4 1 1		1	1 0 1			4 2 1 SE Mst + 2 2 2 A im Y a -	5 6		4	(of)	[trans- tzend- ent]	-al
5	анал- из- а		1 0 Nsr	0 1 2	Nn M g s am a s			1 0 X	5 5	1 1 Mst 2 0 Cpl	(F) g s g	1	5 1 1		1	1 0 1						5		[anal- iz]	

q	Sourc Occur'
6	ос- нов- н- ые
7	теор- ем- ы
8	пс- числ- ен- ия
9	в- ер-о- ост- ей

PART I											PART II																															
											Sec. A.	Section B																											Section C			
																																							Translation			
Source Deriv'ce		Gen. Specfcs. (GS)		Temp. Chs. (TC_j)							Glss. Preds (GP)		TP		Fores't Prds. (FP_{qk})								Sel. Ch. (SC)				L	SP	Hindsight															
				Morph. Specfcs. (MS)							GE & u								Morph. Sps. (MS)				FP						H_1				H_2					H_3					
																													Entry		Resol.		Entry			Res							
		v	c	δ	j	Ps	A	D	tc	C	Y	Z	g	a	d	i	x	n	C	v	P	b	r	k	U	SE	V	A	G	Y	Z	e	Q	K	j	E		C	v	P	f	j	ES	h	Q	K	J	Q	K	j	ES	h	Q	J	R	PI	T	E
	oc- ccu- r- ss	1	1		0 1 2	Aj			ps								1		1	0	X		6 6	1 2	1 0	Cpl Mfr			g M	g	s	1	2	2	2		1	1	0	0													6	(the)	bas- ic	
	reop- ent- n	1	0	Nsr	0 1 2 3	Kn				F			g n a	s p p					1	0	X		7 7	1 2	1 0	Mst Cpl			im	g	a p	1	7	1	3		1	1	0	0													7		[teor- em]	-s
	-c- rcu- ns- ns	1	1		0 1 2 3	Kn	im			K			g n a	s p p			1	1	1	0	X		8 8	1 2	1 0	Cpl Mfr			g im F	a	p	1	8	1	1		1	1	0	0												3	8	(of)	calcul- ation/ us	
	a- ep-o-ti- ocr- ez	1	1	Cpr	0 1	Kn	im			F			g	p			1	1	1	0	X		9 9 25	1 2 3	1 0 3	Cpl Mfr Cpl			g im N	g	s i	1	9	1	1		1	1	0	0													9	(of)	pro- b- ablı- ity	-s

PART I									PART II																
	Source Occur'ce		Gen. Spcfcs. (GS)	Temp. Chs. (TC_j)	Morph. Spcfcs. (MS)	Glss. Preds (GP) GE & u	PS	Sec. A. TP	Section B		Forec't Prds. (FP_{qk})	Morph. Sps. (MS)		Sel. Ch. (SC) FP		L	SP	Hindsight H_1 Entry	H_1 Resol.	H_2 Entry	H_2 Res	H_3	Section C: Translation		
ς		ς	v c δ	j	Ps A D ts G Y Z	g a d i x n		C v P b	r	k U SE	V A G Y Z	e	Q K j	B		C v P	f j ES	h Q K J	Q K j ES	h Q	J	R	PI	T	E
[illegible]	[illegible]	1	1 1	1 2	Cj cd Av cd		cw	1 0 X	10 10 26	1 1 Cpl 2 0 Mfr 3 3 Cpl	g im F g p m	1	0 0 1		1	1 0 0	Ps Cj 2 Ps Av	+ 12 + 12				10		and	
[illegible]	[illegible] rex	1	1 1 Idm	1	Av cd		cw	1 0 X					0 0 1		1	1 0 0						11		at the same time	
[illegible]	[illegible]		1 1	0 1 2	Aj ps n p im e p			1 0 X 1					2 2 2	6	1	1 0 0						12	(the)	main	
13	[illegible]		1 1	0 1 2 3	Nn im N g s n p a p	1		1 0 X	13 13	1 1 Mst 2 0 Cpl	im a p g	1	13 1 3		1	1 0 0						13		ap- plic- ation	-s
14	[illegible]	1	1 1 Spf	1	Pn pn 3 g P			1 0 X	14 14	1 1 Cpl 2 0 Mfr	g im N a p	1	14 1 1		1	1 0 0						14	(of)	them	

PART I							PART II: Sec. A.					Section B								Section C: Translation			
Source Occur'ce	Gen. Spfcs. (GS)	Temp. Chs. (TC_j)	Morph. Spfcs. (MS)	Glss. Preds (GP) GS & u	PS	TP	Fores't Prds. (FP_{qk})		Morph. Sps. (MS)		e	Sel. Ch. (SC) FP	L	SP	Hindsight H_1 Entry	H_1 Resol.	H_2 Entry	H_2 Res	H_3				
	S v|c| S	j	Ps| A| D|ts|G|Y|Z	g|a|d|i|x|n		C|v|P|b	r	k|U| SE	V| A|G|Y|Z			Q|K|j| B		C|v|P	f|j| ES	h| Q|K|J	Q|K|j| ES	h| Q	j	R	PI	T	E
15 [illegible]	0 0				,	2 0 0						0 0 0	1	2 0 0						15			
16 [illegible]	1 1	0 1 2	AJ rl n p im a p		co	2 0 0 1	16 16	1 7 Sbj (2) 7 Pdc				16 1 1 13	1	2 0 0	SE Sbj + 21 2 Y a + 17					16		which	
17 [illegible]	1 1	1	Vb id pr 3 p	2 2 1	vf	2 1 0	17	1 7 Pdc	vf N 3 p			17 1 1	1	2 1 0						17		serve	
18 [illegible]	1 1	0 1	Nn im F i s	2 1 1		2 1 0	19 18	1 5 Cpl 2 5 Cpl	d i	1 1		18 2 1	1	2 1 0						18	(by/ as the)	aid	
19 [illegible]	1 1	0 1 2	AJ ps (F) i s d p			2 1 0	19 19 19	1 1 Cpl 2 0 Mfr 3 5 Cpl 4 3 Cpl	g im F i s d i	1 1 1		19 3 2	1	2 1 0			19 4 1 SE Mst + 20			19	(to/for)	co- p- ious	

q	Source Occur'ce	ø
20	ЗН-ЗН-НЯМ	
21	,	
22	—	
23	ВОТ	
24		
25	ИНОЛ-Ь	

PART I								PART II Sec. A.	Section B									Hindsight					Section C : Translation			
	Source Occur'ce	∅	Gen. Spcfcs. (GS) v|c|s	Tmp. Chs. (TC_j) j	Morph. Spcfcs. (MS) Ps|A|D|ts|C|Y|Z	Glss. Preds (GP) GE & u g|a|d|i|x|n	PS	TP C|v|P|b	r	Fores't Prds. (FP_{qk}) k|U|SF	Morph. Sps. (MS) V|A|C|Y|Z	e	Sel. Ch. (SC) FP Q|K	j	B	L	SP C|v|P	H_1 Entry f|j| ES	H_1 Resol. h|Q|K	H_2 Entry Q|K|i| ES	H_2 Res h|Q	H_3 j	R	PI	T	E
20	[illegible]		1 1	0 1	Nn im N d p	1 1		2 1 0	20 20	1 1 Mst 2 0 Cpl	d p g	1	20 1 1			1	2 1 0						20		know- ledge	-s
21	,		0 0 ,				,	1 0 0		1 3 Cpl	m		0 0 0			1	1 0 0						21		,	
22	–		0 0 –				–	1 0 0					0 0 0			1	1 0 0						22		–	
23	BOT	1	1 1	1	Av dm	1	pc	1 1 0					0 0 1			1	1 1 0						23		here	
24			1 0 Sis	1	Vb cp id pr			1 1 0		1 3 Cpl	d	1	2 1 1			1	1 1 0						24		(is)	
25	[illegible]		1 1	0 1 2	Nn im F n s a s	1 1		1 1 0	25 25	1 5 Cpl 2 4 Cpl	n i		25 1 1			1	1 1 0					2	25	(the)	idea	

PART I								PART II Sec. A.	Section B													Section C			
q	Source Occur'ces	∅	Gen. Specfcs. (GS) v e δ	Temp. Chs. (TC_j) j	Morph. Specfcs. (MS) Ps A D tc G Y Z	Glss. Preds (GP) GE & u g a d i x n	PS	TP C v P b	Forec't Prds. (FP_{qk}) r	k U SK	Morph. Sps. (MS) V A G Y Z	e	Sel. Ch. (SC) FP Q K j B	L	SP C v P	Hindsight H_1 Entry f j ES	H_1 Resol. h Q Kp	H_2 Entry Q K j ES	H_2 Res h Q	H_3 J	Translation R	PI	T	E
26			0 0					1 1 1		1 3 Cpl	m		0 0 0	1	1 1 1						26			
27	ose	1	1 1	0 1 2	Pn pn 1 d s ℓ s			1 1 1					0 0 1	2	1 1 1	1 Y d - 29 1 + 33					28		suggest- ed	
28	upei- nos- ens- ax		1 1	0 1	Aj pv pr F n s	2 2 1 1	po	1 1 1 1					25 1 1 25	2	1 1 1						27	(to/for)	me	
29	rocnod- ne- ox		1 1	0 1	Mn sm tl N i s	1		1 1 1	27 29	1 5 Cpl 2 5 Cpl 3 3 Cpl	d i m	1 1	29 2 1	2	1 1 1						29	(by)	Mr.	
30	C.		1 0 Inl	1 2	Mn pw Np			1 1 1	30 30	1 1 Cpl 2 0 Mfr 3 3 Cpl	g pw M i s d	1 1	30 2 1	2	1 1 1					z	30		[s.]	

PART I											PART II																																					
	Source Occurrce		Gen. Spcfcs. (GS)			Temp. Chs. (TC_j) Morph. Spcfcs. (MS)							Clas. Preds (CP) GE & u					PS	Sec. A. TP					Fores't Prds. (FP_{qk})										Section B Sel. Ch. (SC)				L	SP	Hindsight														Section C Translation				
																																									H_1 Entry			H_1 Resol.			H_2 Entry				H_2 Res		H_3					
q		ø	v	c	δ	j	Ps	A	D	tc	C	Y	Z	g	a	d	i	x	n		C	v	P	b	r	k	U	SE	V	A	G	Y	Z	e	Q	K	j	B		C	v	P	f	j	ES	h	Q	K	J	Q	K	j	ES	h	Q	j	R	PI	T	E
31	Г.		1	0	Inl	1 2	Nn Np	pw													1	1	1		31 31	1 2	1 0	Mfr Cpl	pw		M g	i	s	1	31	1	1		2	1	1	1														2	31		[Q]	
32	С- тror- ан-ов- ки		1	0	Cap	1	Nn	pw													1	1	1		32 32	1 2	1 0	Mfr Cpl	pw		M g	i	s	1	32	1	1		2	1	1	1															32		[S- trog- an-ov]	
33			0	0	Esp													es	0	0	0						1 2	1 0	Cpl Mfr			g M	i	s	1	0 0 0				1	0	0	0															33			

Notes for Chart

Note no.	q	Headings	Explanation
1	all	all	The two Appendices in this paper define all headings and entries in the chart. All terms found therein are capitalized in the following Notes.
2		Occurrence	This portion, although overwritten in Part II, is retained here for the convenience of the reader. It indicates, on separate lines, the 1) Pseudo-prefixes, 2) Pseudo-roots, 3) Pseudo-suffixes, and 4) Ending of a Word-utterance which is listed in the Glossary.
3		∅	This Flag is not part of the routine. It is introduced for the convenience of the reader to indicate a Word-utterance which is found in the Special List.
4		PS	This portion, although overwritten in Part II, is retained for the convenience of the reader. Definitions of its entries will be found in the List of Symbols.
5		r	This signal is not part of the routine. It is introduced for the convenience of the reader to indicate the number of the String which caused the deletion of the Foresight Prediction alongside of which the r appears.
6		FP_{qk}	The predictions entered in String q are made by the String q-1.
7		H_1 to H_3	These entries are made within the String q and are investigated by subsequent Strings.
8		SC and H_3	For convenience of the reader, these entries are placed on the same line with the corresponding TC_j.
9		PI	A subroutine makes these crude Pretarget Inserts. They are printed in parentheses.

TBS 150
Manuscript
Paper

GPO
16—11141-1

Note no.	q	Headings	Explanation
10		T	In the Glossary, English correspondents are stored in a compact form somewhat similar to the Source Transform. The decomposition is shown here. All transliterated portions of T are printed in brackets. Target Inserts are printed in parentheses.
11		E	The English Pidgin Endings assume all words to be regular.
12	1	GP	The numbers (u) in the columns, not headed by x, will give rise to the FP's in String 2. Their connotation is explained under the heading "Signal, urgency" in the List of Terms.
13		$FP_{1,2}$	The r in this Prediction is encircled to indicate (to the reader only) that it is amplified within String 2 in the light of the chosen Subject.
14		SC	The Q.K.j indicate that TC_1 (a verb in the infinitive) was chosen as the Subject.
15		H_1	The unflagged Entries indicate that the SC was a Doubtful Choice. They are to be interpreted as follows: First Entry a. The j is blank to indicate that the Entry concerns the SC. b. The SE indicates that this part of the SC contains the doubtful item. c. The Sbj gives the specific Element of the SC which is doubtful. Second Entry a. The j indicates that TC_1 may be governed, not by $FP_{1,1}$ as chosen, but by some future FP. b. The Y indicates the part of some future FP which may govern TC_1. c. The m indicates that TC_1 may be the fulfillment of a Prediction for a verb in the infinitive mood.

NBS 156
Manuscript
Paper

GPO
10—11141-2

Note no.	q	Headings	Explanation
16	2	TP	The preposition starts a new phrase, as indicated by P = 1.
17		$FP_{2,k}$	The first Prediction is $FP_{1,2}$ revised in the light of the chosen Subject. The number 3 in the Y column indicates the 3rd person. The remaining FP were culled from GP in String 1. The Flag e indicates that a crude PI will be inserted before a T which fulfills a predicted dative or instrumental Complement.
18		SC	The Q and K are zero, because a preposition is an Unpredictable Choice.
19	3	TC_o	The j is set to zero to indicate that the information will be common to all the TC_j listed in String 3.
20		GP	The Grammar Essential g is not used for nouns or adjectives, since the routine always makes the prediction of a Complement in the genitive for these parts of speech. In Russian, adjectives may act as nouns.
21		H_2, H_3	When the Complement of a preposition is fulfilled, the other TC_j are disregarded and not entered in Hindsight.
22	4	Occurrence	The prefix and first suffix are non-slavic.
23		TC:G	The encircled F means Masculine or Neuter.
24		TP and SP	The routine resolves the Uncertainty Signal X, in TP, concerning the phrase number P.
25		FP	The first two Predictions are made by the routine anent the noun in the preceeding String. They were deleted simultaneously because, whether fulfilled or not, FP's with Urgency 1 or 0 are not kept beyond one String. (Cf. Urgency code in List of Terms). The third FP throws some light on the second query in H_1

IBS 156
Manuscript
Paper

GPO
16—11141-4

Note no.	q	Headings	Explanation
			but does not resolve it, as indicated by the negative sign in h.
26		H_2	Two alternate choices are noted here: First conflict. a. The SE indicates the part of some future SC which may throw light on the conflict. b. The Mst indicates the fact that the conflict will be resolved, if a future SC will supply a Master for the current SC. Second conflict. First Entry a. The A indicates the part in a future SC which may throw light on the conflict. b. The im indicates the fact that the conflict will be resolved, if a future SC will supply an inanimate Master for the current SC. Second Entry. a. The Y indicates the part in a future SC which may throw light on the conflict. b. The a indicates the fact that the conflict may be resolved, if no future SC fulfills the expectation of a Complement in the accusative case.
27		E	This is the Pidgin Ending for all non-slavic adjectives.
28	5	TC:G	The Pseudo-suffix ИЗ with a non-slavic root will be taken as denoting a masculine noun. Any required non-slavic feminine or neuter noun with this Pseudo-suffix will be included in the Glossary.
29		FP	These Predictions are supplied by the routine anent the adjective in the preceeding String.

NBS 156 Manuscript Paper

GPO 16—11141-1

Note no.	q	Headings	Explanation
30		SC and H_2 resol.	The Master prediction was satisfied, thus resolving the first Entry in H_2 (Cf. Note 26.) Unfortunately the Occurrence was a non-slavic word and not represented in our Glossary. There was no information about its animation, and therefore the second Entry can not be resolved.
31		H_2 Entry	The Entry of the form 5.2.1, to indicate that $FP_{5,2}$ is also satisfied by TC_1, is omitted for the following reason. When a Master genitive case prediction is satisfied, there is at present no way to resolve this ambiguity, since the considerations of semantics and of context would be involved.
32	6	SP and $FP_{4,3}$	Since none of the Predictions made by the Strings in phrase 1 are fulfilled by the current String, whereas a Prediction in the main clause is, the phrase number P it is set to zero again. The unfulfilled Prediction in Phrase 1, namely $FP_{4,3}$ which bears a low U is deleted.
33		TC_1 and H_1 resol.	Since TC_1 could serve as Subject, the fact is noted in the resolution column of the first Entry in Hindsight H_1 (in String 1.)
34		SC and H_2	The SC throws partial light on the conflict in the H_2 column. (Cf. Note 26.)
35	7	TC:G	The suffix EM with a non-slavic root will be considered as belonging to feminine nouns. (Cf. Note 28.)
36		TC and H_1 resol.	Although TC_2 could serve as the Subject, no note is made of this fact in the resolution of the first Entry in H_1, because it would be the Master of TC_1 in the previous String, which has already been recorded. (Cf. Note 33.)

BS 156
Manuscript
Paper

GPO
16—11141-1

Note no.	q	Headings	Explanation
37	8	TC_2 and H_1 resol.	Partial light is thrown on the first Entry in H_1.
38		TC_3 and H_3	This Choice is pigeonholed.
39		T	There is no way to resolve the polysemy at the present stage of our investigation.
40	9	$FP_{9,3}$	This prediction deletes $FP_{2,5}$, as explained in the List of Terms under the heading "Urgency Code."
41	10	TC and H_1	The iterative scheme in Section A of Part II established the fact that this String did not start a new clause (since the old clause still had the Status zero). The current String could therefore represent either a coordinative conjunction binding two related SC, or an adverb. Since both Choices are Unpredictable, they are recorded as Doubtful. (unflagged Entries in H_1).
42	11	GS:δ	This String represents an Idiom, which constitutes a single Word-utterance.
43	12	b	The Backward Flag was placed by the Profile routine in Section A of Part II. It indicates that, instead of examining the FP's, the routine will scan the previous SC in reverse order to establish a new link in the structural Chain.
44		SC:B	The entry in this column, indicates that the SC in String 6 is conjunctively related to the current String, and the true nature of String 10 is thus established, as indicated by the resolutions in H_1.
45	14	GS:δ	The source word might under certain conditions be treated like a prepositive possessive modifier rendered in English by "their".
46		TC:D	The number 3 indicates the person.

S 156
uscript
aper

GPO

10—11141-1

Note no.	q	Headings	Explanation
47	16	FP	These Predictions were caused by the change in the clause number C of the previous String.
48		b; B; H_1	Similar entries have been explained in Notes 43, 44, and 15.
49	23	TP:v	The Profile routine indicated that the current String must be followed by a New String representing a copulative verb. The Status Flag v was set to 1 in order to effect the proper insertion. (Cf. Note 50.)
50	24	GS:δ	This String was inserted because of the v Flag in the previous String. (Cf. Note 49.)
51	25	FP	These Predictions were made by the routine anent the copulative verb in the preceeding String.
52	27	TC_2	The locative case cannot be considered, since it is governed only by a preposition.
53		L	The Chain is broken, as the routine cannot account for the dative. (The $FP_{24,1}$ cannot be considered, since the two Strings are not in the same phrase.)
54		H_1:f	The Entry is flagged, since the Choice is Unexpected.
55		R	The rearrangement in the target order will be explained in Note 62.
56	28	b; B; R	Cf. Notes 43, 44, 62.
57	29	$FP_{29,1}$	This prediction throws partial light on the H_1 Entry in String 27. The r is indicated in the light of Note 62.
58		PI	Since the noun is animate, only one connotation of the preposition is used.

NBS 156 Manuscript Paper

GPO

16—11141-2

Note no.	q	Headings	Explanation
59	30	GS:δ and TC:A	This part is used in obtaining the signal pw (proper word.)
60	31	FP	The appearance of an initial calls strongly for another proper noun. Hence the inversion of the usual order of Predictions after a noun.
61	32	CS:δ and TC:A	The fact that the capitalized word does not start a sentence causes A to become pw.
62	33		Since no other explanation offered itself for the H_1 resolution in 27, the previous explanation is accepted, the Entry is deleted, and L is reduced by one. This fact also causes the rearrangement of the two neighboring R's. The T's for these two strings reflect the result of this rearrangement.

End of Iteration 1. Since no FP's with U of 6 or 7 remain and the chain number L is equal to 1, no other iteration is necessary. The translation is printed out as indicated in the last three columns of the chart, followed by the tally of unresolved Hindsight Entries.

38 156
nuscript
Paper

GPO
16—21141-1

APPENDIX I

LIST OF TERMS

ADDRESS, extended (α). The locator of an item on an external storage medium. Its form depends on the machine used. On the IBM 704 it consists of (1) type number, (2) block number, (3) serial position of item in the block.

CHAIN. A group of consecutive SC's characterized by the same chain number (L).

CHOICE,

<u>doubtful</u>. A Selected Choice which entails the recording of an unflagged query in Hindsight H_1.

<u>final</u> (FC). The Selected Choice in the last iterative cycle.

<u>predictable</u>. Not belonging to class of Unpredictable Choices.

<u>unexpected</u>. A Predictable Choice which does not agree with any of the Foresight Predictions (FP).

<u>unpredictable</u>. A TC containing one of the following parts of speech: (1) a conjunction (2) an adverb (3) a particle, and (4) a preposition; or else (5) it is a punctuation mark.

<u>selected</u> (SC). A Temporary Choice selected as a link in the Chain during a current iteration.

<u>temporary</u> (TC). A grammatical interpretation of a Source Occurrence.

CODE, urgency. Cf. Urgency

ELEMENT, sentence. Cf. Sentence

NBS 156
Manuscript
Paper

GPO
16—11141-6

ENDING,

source. The true inflectional ending of a word in the source text.

pidgin (E). Regular endings affixed to stems of target correspondents, regardless of correct usage.

ENTRY,

glossary. A complete set of Glossary items corresponding to a Source Stem.

hindsight. A record of an Unexpected String, a Doubtful Choice, or a surplus Temporary Choice. (cf. Hindsight)

ESSENTIAL, grammar (GE). A grammatical form called for by a glossary Prediction (e.g. an accusative called for by a transitive verb). Each type of GE has a separate location in the string reserved for it. A GE is predicted by storing an Urgency Signal in this location.

EXTENDED Address. Cf. Address.

FILE, sorting. The internal space allotted for sorting the Extended Addresses (α).

FLAG. A binary digit (i.e. either a 0- or 1-bit).

Backward (b). A 1-bit alerting the machine to examine the foregoing Selected Choices in order to establish the linkage of the current String.

Chain (f). A 1-bit accompanying a Hindsight Entry in H_1 to record an Unexpected Choice.

Correspondence (c). A bit indicating whether the String contains target correspondents or not.

IBS 156
Manuscript
Paper

GPO
16—11141-8

FLAG. (continued)

Insert (e). A 1-bit accompanying a Foresight Prediction to indicate that a suitable English preposition is to be used as a Pretarget Insert (PI).

Locative (x). A 1-bit in Glossary Predictions to indicate that the locative case is, 1) if the String represents a preposition, one of the cases governed by the preposition; or 2) if the String does not represent a preposition, to be used after the next Positional Preposition encountered in the sentence.

Status (v). A bit indicating whether the predicate of the current clause has turned up or not.

Word (w). A bit indicating whether the String represents a word or not.

GENERAL Specifications. Cf. Specifications

GLOSSARY. The externally stored source-to-target dictionary used by the MT group at NBS. It is stored in a greatly compacted form and contains diacritical material designed to aid in the syntactic--and to a small degree in the semantic--analysis of source sentences.

GRAMMAR Essential. Cf. Essential.

HINDSIGHT. Internal space allotted for storing in

H_1, Entries concerning Unexpected Choices or Doubtful Choices.

H_2, Temporary Choices (TC), other than the Selected one, which fulfill Foresight Predictions (FP).

BS 156
Manuscript
Paper

GPO
10—11141-1

HINDSIGHT (continued)

H_3, Temporary Choices (TC), other than the Selected one, which do not fulfill any of the Foresight Predictions (FP).

IDENTIFICATION Tag. Cf. Tag.

INSERT,

target (TI). A target correspondent incorporated in a New String.

pretarget (PI). A word inserted before a target correspondent.

LIST. Internally stored one-to-one correspondences, yielding for each of the

endings, an address β for each Source Ending, enabling the machine to find, subsequently, the corresponding morphological information in the Table of Endings.

pseudo-prefixes, a 6-bit character, as a substitute for each Pseudo-prefix.

pseudo-roots, an Extended Address α leading to the location of the first externally stored Satellite of each Pseudo-root.

pseudo-suffixes, a 6-bit character, as a substitute for each pseudo-suffix.

symbols, a definition (to be found at end of the paper)

terms, a definition (to be found at end of the paper)

special words, dictionary information. The arguments in this List consist of conjunctions, prepositions, particles, idioms, abbreviations, some adverbs, and words with ambiguous endings.

MORPHOLOGICAL specifications. Cf. Specifications.

OCCURRENCE, source. A combination of one or more characters in a source text. It may represent: (1) a Word-utterance (2) a punctuation mark or a set of such; (3) a symbol or a set of such; (4) a diagram or a set of such; etc.

POSITIONAL preposition. Cf. Preposition.

PREDICTIONS,

foresight (FP), information concerning TC's which are expected to occur somewhere in the sentence under consideration. Such information is derived either from rules of grammar incorporated in the machine instructions, or from Glossary Predictions, or from the Temporary Profile.

glossary (GP), partial information, retrieved from the Glossary or Special List and stored as part of a String, indicating what kinds of TC's are expected to occur somewhere before or after the current SC in the same sentence. A GP is recorded by assigning an Urgency Signal to a Grammar Essential. One String may contain several GP's.

PREPOSITION,

positional. One of a set of Russian prepositions which govern either the accusative or locative case.

PROFILE. The sequence of sets of designations, incorporated in each String of a sentence, which may throw light upon the ranges of its clauses and phrases.

Selected (SP). The Temporary Profile, revised during an iterative cycle in the syntactic integration process.

SS 150
nuscript
Paper

GPO
10—11141-1

Skeleton (PS). The initial stage of a Profile, which bears one signal for each Occurrence, indicating the latter's significance in determining the clauses and phrases of the current sentence.

Temporary (TP). A sequence of sets of four preliminary designators assigning a rough clause (C) and phrase (P) number, as well as a Status (v)--and possible Backward (b)--Flag, to each Occurrence.

PSEUDO-ROOT. That portion (if any) of a word-Occurrence remaining after the Pseudo-prefixes, Ending, and Pseudo-suffixes are stripped off.

PSEUDO-PREFIX. One of a set of combinations of source letters which are frequently found at the beginning of words in the source language.

PSEUDO-SUFFIX. One of a set of combinations of source letters which are frequently found before the Source Ending of words in the source language.

SATELLITE of a Pseudo-root. A Glossary Entry listing the Transform of a Source Stem which contains the Pseudo-root in question.

SENTENCE element (SE). One of the following ingredients of a sentence:

Subject
Predicate
Complement
Modifier
Master
Clause
Phrase.

SIGNAL,

peculiarity (δ). An indicator of some peculiar nature of an Occurrence, e.g. that

it is a capitalized word,

it is an initial,

its root is not listed, etc.

uncertainty (X). Used instead of a clause or phrase number in a Profile, when the determination of that number is not possible.

urgency (u). One of the numbers 0 to 7, indicating the probability of a Glossary Prediction (GP), used to form Urgency Code U in FP, according to the following relation between u and U:

$0 \sim 1$; $1 \sim 3$; $2 \sim 5$; $3 \sim 7$.

A u signal, 4 units higher than the above, indicates an alternate prediction of a clause.

SORTING file. Cf: File.

SPECIFICATIONS,

entry (ES). Signals in H_1 and H_2 to specify the type of query.

general (GS). Designators in a String, consisting of a Word (w) and Correspondence (c) Flags, as well as a Signal of Peculiarity (δ).

morphological (MS). Designators in the TC_j, FP and H_1 of a String, which deal with the grammatical interpretation of the original Occurrence.

STEM. The portion of a word, remaining after the ending is removed.

STRING. The information, replacing the original Occurrence, which is available to the routine during the process of syntactic integration.

New -. A String which is inserted during the process of syntactical integration.

Old -. A String which is available at the beginning of the syntactic process.

TABLE of Endings. A tabulation of the morphological possibilities of each Source Ending.

TAG, identification (t). A serial number attached to each Source Occurrence of a text sentence. It consists of (1) page number, (2) sentence number, and (3) Occurrence number.

TRANSFORM, source (Δ). A contraction representing the Stem of a Source Occurrence.

URGENCY code (U). One of the numbers 0 to 7, connoting the probability of a Foresight Prediction (FP) as follows:

7, must occur sometime

5, very likely to occur sometime

3, may occur sometime. An FP bearing this U is erased by a subsequent FP identical to it

1, will be the next Choice or won't occur at all. It is erased after the next SC.

A code of 6, 4, 2 and 0, indicates the same degree of Urgency as 7, 5, 3 and 1 respectively. Moreover, the even-numbered codes

denote an FP alternate to the last preceeding odd-numbered FP in the same String (e.g. successive U's of 5,4,2 indicate that the second and third predictions are alternates for the first, so that if one of the three occurs, all three could be erased.) An FP with U > 4 is not erased until the end of the iterative cycle, unless it, or one of its alternates, are satisfied. An FP with U = 6 or U = 7, left unsatisfied at the end of a cycle, calls for another iteration.

URGENCY signal - cf. Signal

UTTERANCE, word -. Cf. Word

WORD - utterance (W). One word or a set of consecutive words--in complete or abbreviated form--used as an entity. (e.g. an initial, compound word, an idiom, etc.)

NBS 156
Manuscript
Paper

GPO

16—11141-4

APPENDIX II

LIST OF SYMBOLS

α	Extended Address
A	Heading in TC_j and FP
a	Accusative case
Aj	Adjective, a Ps
am	Animate
Av	Adverb, a Ps
β	Address of argument in Table of Endings
B	Signal in SC to indicate the String where concatenation was established during backward examination
b	Backward Flag
C	For all Strings other than the first in a sentence, C is the clause number in the Profile. For the first string of a sentence, (since all sentences start with clause number 1) we shall use this symbol as a code: 1, declarative sentence 2, interrogative sentence 3, exclamatory sentence, etc. This is possible because the Temporary Profile is obtained as a result of an iterative routine, and the nature of the sentence is known before Part II is undertaken.
c	Correspondence Flag
Cap	Word starts with capital, in δ
cd	Coordinate
Cj	Conjunction, a Ps
Cls	Clause, an SE
co	Clause opener, in PS
cp	Copulative
Cpl	Complement, an SE
Cpr	Compound root, in δ
cw	Coordinate word, in PS
Δ	Source Transform
δ	Signal of Peculiarity
D	Heading in C_j
d	Dative case
dm	Demonstrative
E	Pidgin Ending
e	A 1-bit to indicate that an English preposition is to be used as a PI
ES	Entry Specifications
es	End of sentence, in PS
Esp	End-of-sentence-period, in δ

38 15C
nuscript
Paper

GPO

10—11141-6

F	Feminine gender
f	Chain Flag
FC	Final Choice
FP	Foresight Prediction
G	Gender: M, F, and N. An encircled gender indicates the <u>two other</u> genders
g	Genitive case
GE	Grammar Essential, such as <u>g</u>, <u>a</u>, <u>d</u>, <u>i</u>, <u>x</u>, <u>m</u>
GP	Glossary Predictions
GS	General Specifications
H_1	Storage of queries anent Unexpected and Doubtful Choices
H_2	Storage of alternate predicted TC
H_3	Storage of surplus unpredicted TC
h	(for convenience of the reader only) Signal for type of resolution in Hindsight: minus for partial, plus for complete
i	Instrumental case
id	Indicative mood
Idm	Idiom, in δ
if	Infinitive mood
im	Inanimate
Inl	Initial (a capital letter followed by a period), in δ
J	A number indicating the jth TC in String Q
j	A serial number of TC within a String <u>q</u>
K	A number indicating the kth FP in String Q
k	A serial number of an FP within a String <u>q</u>
L	Chain number, indicating the degree to which the syntactic integration lacks cohesion
ℓ	locative case
M	Masculine gender
m	A Flag, indicating whether or not a verb in the infinitive is predicted.
Mfr	Modifier, an SE
Mst	Master, an SE
N	Neuter gender
n	Nominative case
Nn	Noun, a Ps
Np	Not part of speech
Nsr	Non-slavic root, in δ

156 script per

GPO

10—11141-1

∅	(for convenience of the reader only) A 1-bit to show that the W was found in the Special List
P	Phrase number in a Profile
p	Plural number
pc	Postpositive copulative implied, in PS
Pdc	Predicate, an SE
PI	Pretarget Insert
Pn	Pronoun, a Ps
pn	Personal
po	Phrase opener, in PS
Pp	Preposition, a Ps.
pr	present tense
PS	Profile Skeleton. A set of signals, in addition to punctuation marks, such as co clause opener cw coordinate word es end of sentence po phrase opener pc postpositive copulative implied vf verb finite vi verb infinitive
Ps	Part of speech
ps	Positive degree
pv	Passive voice
pw	Proper word
Q	Running coordinate for q
q	The serial number of the Strings
R	Target order number, indicating sequence in which the English correspondents will be printed out
r	(for convenience of the reader only), indicator of the String in the consideration of which a given FP is deleted. An encircled (r) indicates the String where the FP is revised.
rl	Relative
S_t	Storage space in internal memory allotted to information about Occurrence t
s	Singular number
Sbj	Subject, an SE
SC	Selected Choice
SE	Sentence Element, in an FP
Sis	String inserted by syntax, in δ
SP	Selected Profile
Sp.	Special possessive form, in δ

156 script per

GPO

16—11141-1

The enclosed Supplement to our NBS Report No. 6595, A NEW APPROACH TO THE MECHANICAL SYNTACTIC ANALYSIS OF RUSSIAN, is intended to clarify some points not treated in the Notes for the Chart starting on page 35. The reader may need guidance in the temporal sequence of the various steps in the routine. We shall therefore review this sequence from the start of the process on through the handling of the first String of the sentence. The Notes following the Chart are designed to clarify situations which do not come up in String 1.

The MT Group at the National Bureau of Standards will welcome further suggestions for making the Chart understandable to all workers in this field.

1. The portions headed "Part I" and "T" in Part II list the material obtained prior to the initiation of the syntactic process.

	PART I																			PART II		
	Source Occur'ce		Gen. Spcfcs. (GS)			Temp. Chs. (TC_j)								Clss. Preds (GP) GE & u						Section C		
							Morph. Spcfcs. (MS)													Translation		
q		ø	w	c	δ	j	Ps	A	D	ts	G	Y	Z	g	a	d	i	x	m		T	E
1	По- каз- а- ть		1	1	Cap	1	Vb		if					7	1	1	1			vi	de- mon- str- ate	

Heading	Explanation
Occurrence	This portion, although overwritten in Part II, is retained here for the convenience of the reader. It indicates, on separate lines, the 1) Pseudo-prefixes, 2) Pseudo-roots, 3) Pseudo-suffixes, and 4) Ending of a Word-utterance which is listed in the Glossary.
ø	This Flag is not part of the routine. It is introduced for the convenience of the reader to indicate a Word-utterance which is found in the Special List.
GS	The General Specifications of this Occurrence indicate that 1) it is a Word-utterance, 2) its correspondent (T) is an English word retrieved from storage, and 3) it starts with a capital letter. (Cf. the headings in the List of Symbols).
TC_j:j	The index j remains at 1; this indicates that the given Occurrence has given rise to only one Temporary Choice.
TC_j:MS	The Morphological Specifications indicate that the Occurrence represents a verb (Vb) in the infinitive mood (if).
GP	The Glossary Predictions accompany the Stem of the Occurrence in Storage. The numbers in the various columns are Urgency Signals (u). Each of these, with the exception of the number headed by x, will give rise to Foresight Predictions to be stored in String 2. Their connotation is explained in the List of Terms, and will be clarified at the time they will be utilized.

Heading	Explanation
PS	The Profile Skeleton, although overwritten in Part II, is kept in the chart for the convenience of the reader. It serves as a basis for determining the boundaries of the clauses and phrases within the source sentence.
Section C T	In the Glossary, English correspondents are stored in compact form somewhat similar to the Source Transform. The decomposition is hown here.

2. The portion headed "TP" lists the components of the Temporary Profile, resulting from the iterative process performed on the data in the Profile Skeleton (PS) involving the entire sentence.

Part II				
Sec. A.				
TP				
q	C	v	P	b
I	1	0	0	

Heading	Explanation
C	The number 1 indicates that the given Occurrence forms part of the first clause
v	This flag remains at the zero level until the Prediction of a predicate is fulfilled
P	The zero indicates that the given Occurrence is not part of a phrase.

3. The following preliminary steps are executed by the routine prior to entering the (minor) iterative cycle involving the first String.

a. The Chain number (L) is set to 1.
b. The Hindsight is cleared of all content.
c. The Foresight Predictions (FP_{qk}) stored within String q are made by String q-1. When q is equal to 1, $FP_{1.1}$ and $FP_{1.2}$ are made by the routine in response to signals indicating the start of the first clause. They predict with utmost assurance (shown by U=7) that a Subject and Predicate will occur within the clause.
d. The Selected Profile (SP) is temporarily set equal to the Temporary Profile.

4. The portion headed "Section B" indicates how the routine attempts to incorporate the given Occurrence as a link of a unifying Chain, representing the sentence structure.

PART II																															
Section B																															
		Fores't Prds. (FP_{qk})									Sel. Ch. (SC)				L	SP			Hindsight												
					Morph. Sps. (MS)														H_1					H_2					H_3	
											FP								Entry			Resol.			Entry			Res		
q	r	k	U	SE	V	A	G	Y	Z	e	Q	K	j	B		C	v	P	f	j	ES	h	Q	K	J	Q	K	j	ES	h	Q	j
1	1	1	7	Sbj							1	1	1		1	1	0	0			SE Sbj	-	6		1						
	(2)	2	7	Pdc																		-	8		2						
																			1	Y	m	-	4	3							
																						-	10	3							
																						-	26	1							
2	24	1	7	Pdc	vf		N	3	s																						
	6	2	7	Cpl				a																							
	6	3	6	Cls				a																							
	24	4	3	Cpl				d		1																					
	9	5	3	Cpl				i		1																					

Heading	Explanation
SC:FP	The Q and K under this heading indicate that $FP_{1.1}$, calling for a Subject, has been responsible for the Selected Choice.
SC:j	The index 1 indicates that TC_1--a verb in the infinitive--fulfilling the above prediction, was selected.
L	The Chain number remains unchanged because the SC fulfilled a Foresight Prediction.
r	This signal is not part of the routine. It is introduced for the convenience of the reader to indicate the number of the String which caused the deletion of the Foresight Prediction alongside of which the r appears. The r accompanying $FP_{1,2}$ is encircled to indicate (to the reader only) that the prediction is amplified within String 2 in the light of the chosen Subject.

Heading	Explanation
$FP_{2,k}$	Foresight Predictions, to be stored in String 2, are made on the basis of the Selected Choice in String 1. First, and most urgent is the amplification of the Predeiction of a Predicate. There were no general grammar Predictions because in our routine a verb does not yield such. The remaining four Predictions came from the Glossary information, accompanying the Temporary Choice selected as SC for String 1. They predict that with utmost necessity a direct object (in the form of either a single word in the accusative case or a clause) will occur and that a slight probability exists for the appearance of complements (i.e. indirect objects) in the dative and/or instrumental case, each followed by a suitable preposition (e Flag = 1). Since the x Flag in String 1 is unity, a suitable signal will be set to indicate that the next Positional Preposition will govern the locative case, unless the signal is turned off by an <u>x</u> equal to zero in a subsequent String.
Hindsight resol.	In general, the TC_j, of the Current String, and the newly made $FP_{p+1,k}$, stored in the next String, are examined to ascertain whether they throw any light on the perplexities, doubts, and conflicts recorded in the Hindsight columns. For q=1, however, the Hindsight is empty at this stage of the process. (The resolutions indicated on the chart for String 1 will, of course, have been made by subsequent Strings.)
SP	In general, the doubts recorded in the selection of the components of the Profile are resolved in light of the SC and of the Hindsight resolution. For q=1, this is seldom necessary.
Hindsight Entries	This information is derived by, and stored in, String <u>q</u> to be investigated by subsequent Strings.

Heading	Explanation
H_1:Entries	The unflagged Entries in String 1 indicate that the SC was a Doubtful Choice. They are to be interpreted as follows: First Entry a. The j is blank to indicate that the Entry concerns the SC. b. The SE indicates that this part of the SC contains the doubtful item. c. The Sbj gives the specific Element of the SC which is doubtful. Second Entry a. The j indicates that TC_1 may be governed, not by $FP_{1,1}$ as chosen, but by some future FP. b. The Y indicates the part of some future FP which may govern TC_1. c. The m indicates that TC_1 may be the fulfillment of a Prediction for a verb in the infinitive mood.
H_2 Entries	The one Temporary Choice in String 1, necessarily taken as the SC, fulfilled (doubtfully) $F_{1.1}$ but not $F_{1.2}$. Therefore, H_2 has no Entry for String 1.
H_3 Entries	When a String has only one TC, there can be no unused Choices.

5. The portion headed "Section C" indicates the order and form in which the target correspondents will appear in the translated sentence.

	Section C			
	Translation			
q	R	PI	T	E
1	1	(To)	de- mon- str- ate	

Heading	Explanation
R	The order in which the correspondents will appear in the printed translation in general follows closely the order of the original Occurrences. On occasion, a particular Selected Choice (SC) and/or a Hindsight resolution may effect a deviation from the sequence. The routine in Section B indicates the proper order (R) of every correspondent in the printed sequence.
PI	A subroutine makes these crude Pretarget Inserts. They are printed in parentheses.
E	The Pidgin Endings assume all English words to be regular.

T	Target correspondent
t	Identification Tag of Occurrence
TC	Temporary Choice
TI	Target Insert
tl	Title
TP	Temporary Profile
ts	Tense
U	Urgency Code, in FP
u	Urgency Signal, in TC
V	Heading in FP
v	Status Flag
Vb	Verb, a Ps
vi	Verb infinitive, in PS
vf	Verb finite, in PS and under heading V in FP
W	Word-utterance
w	Word Flag
X	Signal of Uncertainty, in PS
x	A 1-bit to indicate that the locative case is governed either at once, if the TC represents a preposition, or after the next Positional Preposition (which Cf. in List of Terms.)
Y	A heading in TC and FP
y	Count of Pseudo-prefixes in a word-Occurrence
Z	A heading indicating grammar plurality
z	Count of Pseudo-suffixes in a word-Occurrence.

NBS 156
Manuscript
Paper

U.S. DEPARTMENT OF COMMERCE
Frederick H. Mueller, *Secretary*

NATIONAL BUREAU OF STANDARDS
A. V. Astin, *Director*

THE NATIONAL BUREAU OF STANDARDS

The scope of activities of the National Bureau of Standards at its major laboratories in Washington, D.C., and Boulder, Colo., is suggested in the following listing of the divisions and sections engaged in technical work. In general, each section carries out specialized research, development, and engineering in the field indicated by its title. A brief description of the activities, and of the resultant publications, appears on the inside of the front cover.

WASHINGTON, D.C.

ELECTRICITY. Resistance and Reactance. Electrochemistry. Electrical Instruments. Magnetic Measurements. Dielectrics.

METROLOGY. Photometry and Colorimetry. Refractometry. Photographic Research. Length. Engineering Metrology. Mass and Scale. Volumetry and Densimetry.

HEAT. Temperature Physics. Heat Measurements. Cryogenic Physics. Rheology. Molecular Kinetics. Free Radicals Research. Equation of State. Statistical Physics. Molecular Spectroscopy.

RADIATION PHYSICS. X-Ray. Radioactivity. Radiation Theory. High Energy Radiation. Radiological Equipment. Nucleonic Instrumentation. Neutron Physics.

CHEMISTRY. Surface Chemistry. Organic Chemistry. Analytical Chemistry. Inorganic Chemistry. Electrodeposition. Molecular Structure and Properties of Gases. Physical Chemistry. Thermochemistry. Spectrochemistry. Pure Substances.

MECHANICS. Sound. Pressure and Vacuum. Fluid Mechanics. Engineering Mechanics. Combustion Controls.

ORGANIC AND FIBROUS MATERIALS. Rubber. Textiles. Paper. Leather. Testing and Specifications. Polymer Structure. Plastics. Dental Research.

METALLURGY. Thermal Metallurgy. Chemical Metallurgy. Mechanical Metallurgy. Corrosion. Metal Physics.

MINERAL PRODUCTS. Engineering Ceramics. Glass. Refractories. Enameled Metals. Constitution and Microstructure.

BUILDING RESEARCH. Structural Engineering. Fire Research. Mechanical Systems. Organic Building Materials. Codes and Safety Standards. Heat Transfer. Inorganic Building Materials.

APPLIED MATHEMATICS. Numerical Analysis. Computation. Statistical Engineering. Mathematical Physics.

DATA PROCESSING SYSTEMS. Components and Techniques. Digital Circuitry. Digital Systems. Analog Systems. Applications Engineering.

ATOMIC PHYSICS. Spectroscopy. Radiometry. Mass Spectrometry. Solid State Physics. Electron Physics. Atomic Physics.

INSTRUMENTATION. Engineering Electronics. Electron Devices. Electronic Instrumentation. Mechanical Instruments. Basic Instrumentation.

Office of Weights and Measures.

BOULDER, COLO.

CRYOGENIC ENGINEERING. Cryogenic Equipment. Cryogenic Processes. Properties of Materials. Gas Liquefaction.

IONOSPHERE RESEARCH AND PROPAGATION. Low Frequency and Very Low Frequency Research. Ionosphere Research. Prediction Services. Sun-Earth Relationships. Field Engineering. Radio Warning Services.

RADIO PROPAGATION ENGINEERING. Data Reduction Instrumentation. Radio Noise. Tropospheric Measurements. Tropospheric Analysis. Propagation-Terrain Effects. Radio-Meteorology. Lower Atmosphere Physics.

RADIO STANDARDS. High frequency Electrical Standards. Radio Broadcast Service. Radio and Microwave Materials. Atomic Frequency and Time Standards. Electronic Calibration Center. Millimeter-Wave Research. Microwave Circuit Standards.

RADIO SYSTEMS. High Frequency and Very High Frequency Research. Modulation Research. Antenna Research. Navigation Systems. Space Telecommunications.

UPPER ATMOSPHERE AND SPACE PHYSICS. Upper Atmosphere and Plasma Physics. Ionosphere and Exosphere Scatter. Airglow and Aurora. Ionospheric Radio Astronomy.

NBS